Λ-MINIMAX ESTIMATION II: THE DESIGN PROBLEM

BU-279-M Daniel L. Solomon June, 1970

Abstract

It is suggested that in the presence of incomplete prior information the decision maker may be able to improve his prior knowledge, and that this possibility be included in the cost structure of the decision problem. This gives rise to a design problem of selecting not only an optimal experiment (e.g. sample size) but also an optimal specification of improved prior information. The model is illustrated by determining optimal designs for two Λ-minimax estimation problems, for which the terminal action problems have been considered in an earlier paper.

Biometrics Unit Mimeo Series, Cornell University, Ithaca, New York 14850.

Λ-MINIMAX ESTIMATION II: THE DESIGN PROBLEM

Daniel L. Solomon

It is suggested that in the presence of incomplete prior information the decision maker may be able to improve his prior knowledge, and that this possibility be included in the cost structure of the decision problem. This gives rise to a design problem of selecting not only an optimal experiment (e.g. sample size) but also an optimal specification of improved prior information. The model is illustrated by determining optimal designs for two Λ-minimax estimation problems, for which the terminal action problems have been considered in an earlier paper [1].

1. INTRODUCTION

In the situation in which the decision maker has prior information but cannot completely specify a prior distribution, it has been suggested by several authors (see references in [1]) that he apply the Λ-minimax method. That is, suppose that although the decision maker cannot specify his prior distribution λ, that he can assert that $\lambda \in \Lambda$, where Λ is some subset (called an incompleteness specification) of the class of all prior distributions on the states of nature. A rule $\delta \in D$ which minimizes $\sup_{\lambda\in\Lambda} r^*(\lambda,\delta)$ is called Λ-minimax in D. Here $r^*(\lambda,\delta)$ is the Bayes risk of δ with respect to the prior distribution λ, and D is a set of decision rules under consideration. If Λ contains only one distribution, then a Λ-minimax rule is a Bayes rule with respect to that distribution, while if Λ is the class of all distributions on the states of nature, then a Λ-minimax rule is a minimax rule.

It is now supposed that the decision maker can improve his prior information (that is, further restrict Λ) by for example, introspection, consultation or an interviewing technique such as that due to Winkler [2]. This possibility is to be included in the cost structure of the model. If we assume that the various costs are additive, then the loss function for the decision problem is

$$L(\Lambda,e,\delta,x,\theta) = \alpha(\Lambda) + c_s(e) + \ell(\delta(x),\theta) , \tag{1}$$

where $\alpha(\Lambda)$ is the cost associated with obtaining the incompleteness specification Λ, $c_s(e)$ is the cost of performing experiment e, and $\ell(\delta(x),\theta)$ is the loss associated with observing the datum x, and taking terminal action $\delta(x)$ when θ is the state of nature. The terminal action problem of determining a Λ-minimax rule δ, has been dealt with for some estimation problems in [1]. This paper illustrates the corresponding design problems. The design problem consists of selecting Λ and e to minimize the Λ-minimax expected loss, that is to minimize

$$\alpha(\Lambda) + c_s(e) + \inf_{\delta\in D}\sup_{\lambda\in\Lambda} E\ell(\delta(X),\underline{\theta}) .$$

(The expectation is with respect to the joint distribution of X and θ.)

2. EXAMPLE 1: ESTIMATION OF A LOCATION PARAMETER

Suppose that X and $\underline{\bar{\theta}}$ are random variables (not necessarily Normal), that $n > 0$ and $\sigma_1^2 > 0$ are known real constants and that

$$E(X|\underline{\bar{\theta}}) = \underline{\bar{\theta}} , \qquad \mathrm{var}(X|\underline{\bar{\theta}}) = n^{-1}\sigma_1^2 .$$

(It is convenient to view X as the sample mean of n independent observations on a random variable with (conditional) variance σ_1^2.)

Let

$$D_L = \{\delta|\delta(x) = bx + c\}$$

be the class of all real valued linear functions of a real variable so that for $\delta \in D_L$, $\delta(X)$ is a linear "estimator" of the location "parameter" $\bar{\underset{\sim}{\theta}}$. Suppose that if X is observed to have the value x, and $\bar{\underset{\sim}{\theta}}$ is estimated by $\delta(x)$ when $\bar{\underset{\sim}{\theta}} = \theta$, then the loss incurred is $\ell(\delta(x),\theta) = (\delta(x) - \theta)^2$.

Suppose that the decision maker's prior knowledge is such that his prior distribution λ for $\bar{\underset{\sim}{\theta}}$ has

$$E_\lambda(\bar{\underset{\sim}{\theta}}) = \mu \quad \text{and} \quad var_\lambda(\bar{\underset{\sim}{\theta}}) = \sigma_o^2 , \tag{2}$$

where $\sigma_o^2 > 0$ is known. Further suppose that although he cannot specify μ, the decision maker can learn at cost

$$\alpha(M) = \frac{\alpha^2}{M^2} , \quad \alpha \geq 0, \ M \geq 0 ,$$

that $\mu \in U = \{\mu \,\big|\, |\Delta - \mu| \leq M\}$, so that Λ is the class of all distributions satisfying (2), with $\mu \in U$. Finally, if the cost of observing X is proportional to n (i.e., sampling cost is proportional to sample size), then

$$c_s(e) = c_s n, \quad c_s \geq 0 .$$

Assuming additivity of the various costs as in (1), the overall loss function for this problem is

$$L(M,n,\delta,x,\theta) = \frac{\alpha^2}{M^2} + c_s n + (\delta(x) - \theta)^2 .$$

Thus the decision maker selects M and n, pays $\frac{\alpha^2}{M^2} + c_s n$, receives Δ and x, takes action $\delta(x)$ and incurs the loss $(\delta(x) - \theta)^2$ if θ is the state of nature. By corollary 12.1 in [1], the Λ-minimax expected loss of the Λ-minimax rule (in D_L) is

$$L^*(M,n) \equiv \inf_{\delta \in D_L} \sup_{\mu \in U} E\,L(M,n,\delta,X,\bar{\underset{\sim}{\theta}}) = \frac{\alpha^2}{M^2} + c_s n + \left[\frac{n}{\sigma_1^2} + \frac{1}{\sigma_o^2 + M^2} \right]^{-1} .$$

The design problem consists of finding M and n to make this a minimum.

Since $\frac{\partial^2 L^*(M,n)}{\partial^2 n^2} > 0$ for $n \geq 0$, n may be treated as a continuous variable in seeking the minimum. If the minimizing value, n_o, is not an integer, the optimal size is whichever of $[n_o]$ and $[n_o] + 1$ makes L^* smaller. Let

$$S = \{(\sigma_o, \sigma_1, c_s, \alpha) \mid \sigma_o > 0,\ \sigma_1 > 0,\ c_s > 0,\ \alpha > 0\}\ ,$$

$$\rho = (\sigma_1\sqrt{c_s} - \alpha)/\sigma_o^2\ ,$$

and $\{S_1, S_2, S_3\}$ be the partition of S defined by

$$S_1 = \{(\sigma_o, \sigma_1, c_s, \alpha) \mid 0 < \rho \leq 1\}$$

$$S_2 = \{(\sigma_o, \sigma_1, c_s, \alpha) \mid \rho > 1\}$$

$$S_3 = \{(\sigma_o, \sigma_1, c_s, \alpha) \mid \rho \leq 0\}\ .$$

<u>Theorem 1</u>: With notation as above,

$$L^* \equiv \inf_{M,n} L^*(M,n) = \begin{cases} 2\sigma_1\sqrt{c_s} - \sigma_o^2\rho^2 & \text{on } S_1 \\ 2\alpha + \sigma_o^2 & \text{on } S_2 \\ 2\sigma_1\sqrt{c_s} & \text{on } S_3\ , \end{cases}$$

and the infimum is attained at

$$n_o = \begin{cases} \sigma_1(1 - \rho)/\sqrt{c_s} & \text{on } S_1 \\ 0 & \text{on } S_2 \\ \sigma_1/\sqrt{c_s} & \text{on } S_3\ , \end{cases}$$

$$M_o^2 = \begin{cases} \alpha/\rho & \text{on } S_1 \\ \alpha & \text{on } S_2 \\ \infty & \text{on } S_3\ . \end{cases}$$

The Λ-minimax rule in D_L is then

$$\delta_o(x) = \begin{cases} (1 - \rho)x + \rho\Delta & \text{on } S_1 \\ \Delta & \text{on } S_2 \\ x & \text{on } S_3 \end{cases}$$

A proof is given in the appendix.

A few remarks about the regions S_2 and S_3 are in order. First note that ρ can be viewed as a "prior preference" parameter. That is, ρ is an increasing function of the sampling variance σ_1^2, and the sample size cost parameter c_s, and a decreasing function of the prior specification cost parameter α and the prior variance σ_o^2. Thus in S_2, with ρ large, sampling is imprecise or costly, or it is inexpensive to improve the specification of the prior distribution or the prior distribution is precise. Thus it is appropriate not to sample ($n_o = 0$) and to rely entirely on the prior distribution. In S_3 with ρ small, the converse is true, and it is appropriate to ignore the prior ($M_o^2 = \infty$). The corresponding decision rules are as one would expect.

<u>Corollary 1.1</u>: With notation as in theorem 1, the sample size required for the Λ-minimax procedure does not exceed the sample size required for the procedure with $\delta(x) = x$.

<u>Proof</u>: For the procedure with rule $\delta(x) = x$, the risk is

$$c_s n + E\{(X - \theta)^2 | \theta\} = c_s n + \frac{\sigma_1^2}{n} .$$

Minimizing on n, the optimal sample size n_1, is

$$n_1 = \frac{\sigma_1}{\sqrt{c_s}} .$$

From theorem 1 in each of S_1, S_2, and S_3

$$n_o = \frac{\sigma_1}{\sqrt{c_s}}\left[1 - \frac{\alpha}{M_o^2}\right] \leq n_1 .$$

3. EXAMPLE 2: ESTIMATION OF A SCALE PARAMETER

Let $X_1, X_2, \cdots, X_n$ be independent, identically Normally distributed random variables with mean θ and variance h^{-1}. Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $V = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, where $V = 0$ if $n \leq 1$. The joint density of $\bar{X}$ and V is

$$f(\bar{x},v|\theta,h) = (\text{const.})\left[e^{-\frac{1}{2}hn(\bar{x}-\theta)^2}h^{\frac{1}{2}}\right]\left[e^{-\frac{1}{2}hrv}h^{\frac{1}{2}r}\right]$$

where $r = n - 1$.

Suppose that h^{-1} is to be estimated by a linear function δ of the sufficient statistic, V, with loss function

$$\ell(\delta(v),h) = h^2(\delta(v) - h^{-1})^2$$

so that a misestimation of small values of h^{-1} is more costly than an equal misestimation of large values.

It is now supposed that the family of natural conjugate prior distributions is rich enough to include an incompleteness specification acceptable to the decision maker. The family of natural conjugate prior distributions for $\tilde{\theta}$ and $\tilde{h}$ is the Normal-gamma with density

$$\begin{aligned} f_{N_\gamma}(\theta,h|\mu,v',n^*,n') &= f_N(\theta|\mu,hn^*)f_{\gamma_2}(h|v',n') \\ &= (\text{const.})\left[e^{-\frac{1}{2}hn^*(\theta-\mu)^2}h^{\frac{1}{2}}\right]\left[e^{-\frac{1}{2}hn'v'}h^{\frac{1}{2}n'-1}\right] \end{aligned} \tag{3}$$

for $-\infty \le \theta \le \infty$, $h \ge 0$, and v', n^*, $n' > 0$. If prior knowledge about v' is incomplete and the decision maker has learned

$$v' \in U = \{v' \mid \frac{\Delta}{A} \le v' \le \Delta A\} ,$$

where $\Delta \ge 0$, $A \ge 1$, then Λ is the set of Normal-gamma distributions (3), with $v' \in U$.

Restricting the discussion to rules linear in V, write

$$\delta(v) = bv + c$$

and a Λ-minimax rule requires b,c to minimize $\sup_{v' \in U} E\tilde{h}^2(\delta(V) - \tilde{h}^{-1})^2$. It has been shown in [1, section 5.2] that the Λ-minimax rule is $\delta_o(v) = b_o v + c_o$ where

$$b_o = \frac{\frac{2 + n'G^2}{n' + 2}}{\frac{2}{r} + \frac{2 + n'G^2}{n' + 2}} , \qquad G = \frac{A^2 - 1}{A^2 + 1} , \qquad r = n - 1,$$

and

$$c_o = 2\Delta \frac{A}{A^2 + 1} \frac{n'}{n' + 2} (1 - b_o) .$$

Furthermore, the Λ-minimax risk of δ_o is

$$\sup_{v' \in U} E\tilde{h}^2(\delta_o(V) - \tilde{h}^{-1})^2 = \frac{2(2 + n'G^2)}{2(n' + 2) + r(2 + n'G^2)} .$$

The design problem is that of determining n, the sample size and A, the determinant of the incompleteness specification, U. Postulate a cost $\varphi(A)$ of learning $\frac{\Delta}{A} \le v' \le \Delta A$ and a cost $c_s r = c_s(n - 1)$, $c_s \ge 0$, of obtaining a sample of size n. (This choice of sampling cost gives the same minimizing n as $c_s \cdot n$ but simplifies the computations.) Again assume that all costs are additive so that the design problem is to find r and A to minimize

$$\varphi(A) + c_s r + \frac{2(2 + n'G^2)}{2(n' + 2) + r(2 + n'G^2)} \; .$$

For this example a convenient choice for φ is to set $t = \frac{2}{n' + 2}$, $\bar{t} = 1 - t$, and $y = t + \bar{t}G^2$, and choose

$$\varphi(A) = \frac{\varphi}{y^2} , \qquad \varphi \geq 0 \; .$$

The loss function becomes

$$L^*(y,r) = \frac{\varphi}{y^2} + c_s r + \frac{2y}{2 + ry} \; ,$$

and this is to be minimized over values of y and r such that $t \leq y \leq 1$, $r \geq -1$. As in section 2, r may be treated as continuous. Let

$$S = \{(t,\varphi,c_s) | 0 \leq t \leq 1, \; \varphi \geq 0, \text{ and } c_s \geq 0\} \; ,$$

$$S_1 = \{(t,\varphi,c_s) \in S \; | t \leq \frac{\varphi}{c_s} \leq 1\}$$

$$S_2 = \{(t,\varphi,c_s) \in S \; | \; \frac{\varphi}{c_s} < t\}$$

$$S_3 = \{(t,\varphi,c_s) \in S \; | \; \frac{\varphi}{c_s} > 1\}$$

$$T_1 = \{(t,\varphi,c_s) \in S \; \Big| \sqrt{\frac{2}{c_s}} - \frac{2c_s}{\varphi} \geq -1\}$$

$$T_2 = \{(t,\varphi,c_s) \in S \; \Big| \sqrt{\frac{2}{c_s}} - \frac{2}{t} \geq -1\}$$

$$T_3 = \{(t,\varphi,c_s) \in S \; \Big| \sqrt{\frac{2}{c_s}} - 2 \geq -1\} = \{c_s | c_s \leq 2\} \; .$$

Finally, let $\bar{T}_j$ be the complement (in S) of T_j, $j = 1, 2, 3$, and note that $S = S_1 \cup S_2 \cup S_3$.

Theorem 2: With notation as above

$$\inf_{y,r} L^*(y,r) = \begin{cases} 2\sqrt{2c_s} - \dfrac{c_s^2}{\varphi} & \text{on } S_1 \cap T_1 \\[2ex] \dfrac{c_s^2}{\varphi} + \dfrac{2\varphi}{2c_s - \varphi} - c_s & \text{on } S_1 \cap \bar{T}_1 \\[2ex] \dfrac{\varphi}{t^2} + 2\sqrt{2c_s} - \dfrac{2c_s}{t} & \text{on } S_2 \cap T_2 \\[2ex] \dfrac{\varphi}{t^2} + \dfrac{2t}{2-t} - c_s & \text{on } S_2 \cap \bar{T}_2 \\[2ex] \varphi + 2\sqrt{2c_s} - 2c_s & \text{on } S_3 \cap T_3 \\[2ex] \varphi + 2 - c_s & \text{on } S_3 \cap \bar{T}_3 \end{cases}$$

and is attained at

$$r_o = \begin{cases} \sqrt{\dfrac{2}{c_s}} - \dfrac{2c_s}{\varphi} & \text{on } S_1 \cap T_1 \\[2ex] \sqrt{\dfrac{2}{c_s}} - \dfrac{2}{t} & \text{on } S_2 \cap T_2 \\[2ex] \sqrt{\dfrac{2}{c_s}} - 2 & \text{on } S_3 \cap T_3 \\[2ex] -1 & \text{elsewhere on } S \end{cases}$$

and

$$y_o = \begin{cases} \dfrac{\varphi}{c_s} & \text{on } S_1 \\[2ex] t \ (A = 1) & \text{on } S_2 \\[2ex] 1 \ (A = \infty) & \text{on } S_3 \ . \end{cases}$$

A proof is given in the appendix.

Conclusions similar to those of the previous section can be drawn concerning the optimal y and r in the several regions. Recall that $y = t$ implies that $A = 1$ and $y = 1$ implies $A = \infty$.

A final observation is appropriate.

Corollary 2.1: The optimal sample size for the Λ-minimax estimate does not exceed the optimal sample size for the estimate $\delta(v) = v$, where v is an observed value of

$$V = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 .$$

Proof: Note that for the Λ-minimax procedure

$$r_o = \begin{cases} \sqrt{\frac{2}{c_s}} - \frac{2}{y_o} & \text{on } S_1 \cap T_1,\ S_2 \cap T_2,\ S_3 \cap T_3 \\ -1 & \text{elsewhere on } S, \end{cases}$$

and that $y_o \geq 0$.

For the procedure $\delta(v) = v$, the total risk is

$$c_s n + Eh^2(V - \frac{1}{h})^2 = c_s n + \frac{2}{n-1} ,$$

(see [1, (31)]). The minimizing n is

$$n_1 = 1 + \sqrt{\frac{2}{c_s}} ,$$

or equivalently

$$r_1 = n_1 - 1 = \sqrt{\frac{2}{c_s}} .$$

It follows that $r_o \leq r_1$, and the corollary is proved.

REFERENCES

[1] Solomon, D. L. "Λ-minimax estimation I: two terminal action problems," submitted.

[2] Winkler, R. L. "The assessment of prior distributions in Bayesian analysis," Journal of the American Statistical Association, 62 (1967), 776-800.

APPENDIX OF MATHEMATICAL PROOFS

Theorem 1:

Proof: M and n are sought to minimize

$$L^*(M,n) = \frac{\alpha^2}{M^2} + c_s n + \frac{\sigma_1^2(\sigma_o^2 + M^2)}{\sigma_1^2 + n(\sigma_o^2 + M^2)} \quad .$$

Setting the partial derivatives with respect to M and n equal to zero, it follows after some simplification, that the critical equations are

$$M^2 = \frac{\alpha(n\sigma_o^2 + \sigma_1^2)}{\sigma_1^2 - \alpha n} , \quad n < \frac{\sigma_1^2}{\alpha} \tag{4}$$

and

$$(\sqrt{c_s}\, n - \sigma_1)(\sigma_o^2 + M^2) + \sigma_1^2\sqrt{c_s} = 0 \; . \tag{5}$$

Substituting (4) into (5) gives

$$n_o = \sigma_1(1 - \rho)/\sqrt{c_s} \; , \qquad \rho \leq 1 \tag{6}$$

where

$$\rho = (\sigma_1\sqrt{c_s} - \alpha)/\sigma_o^2 \; ,$$

and substituting n_o back in (4) gives

$$M_o^2 = \alpha/\rho \; . \tag{7}$$

With some computation, it can be shown that $n_o < \frac{\sigma_1^2}{\alpha}$ if and only if $\rho > 0$, and that the matrix of second derivatives is positive definite so that the solution is in fact a minimum. So in the case $0 < \rho \leq 1$; that is, in S_1, a solution is given by

(6) and (7). By a continuity argument, take $n_o = 0$ if $\rho > 1$ [see display (6)]; that is, in S_2. In this case, by substitution in (4) it follows that $M_o^2 = \alpha$. Finally, if $\rho \leq 0$ (in S_3), take $M_o^2 = \infty$ [see display (7)], and so $L^*(\infty,n) = c_s n + \frac{\sigma_1^2}{n}$, which is minimized for $n_o = \frac{\sigma_1}{\sqrt{c_s}}$.

The proof is completed by substitution of n_o and M_o^2 into $L^*(M,n)$ to compute L^* and into

$$\delta_o(x) = \frac{\dfrac{nx}{\sigma_1^2} + \dfrac{\Delta}{\sigma_o^2 + M^2}}{\dfrac{n}{\sigma_1^2} + \dfrac{1}{\sigma_o^2 + M^2}} .$$

Theorem 2:

Proof: Differentiating $L^*(y,r)$ with respect to y and r and setting the derivatives equal to zero gives

$$\frac{y_o^3}{(2 + r_o y_o)^2} = \frac{\varphi}{2} , \tag{8}$$

and

$$\frac{y_o^2}{(2 + r_o y_o)^2} = \frac{c_s}{2} . \tag{9}$$

Substituting (8) into (9) gives

$$y_o \frac{c_s}{2} = \frac{\varphi}{2} ,$$

and thus

$$y_o = \frac{\varphi}{c_s} . \tag{10}$$

Substituting (10) into (9) gives

$$\frac{\frac{\varphi}{c_s}}{2 + r_o \frac{\varphi}{c_s}} = \sqrt{\frac{c_s}{2}}$$

and thus

$$r_o = \frac{\varphi - 2\sqrt{\frac{c_s}{2}}\, c_s}{\varphi\sqrt{\frac{c_s}{2}}} = \sqrt{\frac{2}{c_s}} - 2\,\frac{c_s}{\varphi} \tag{11}$$

With some computation it can be shown that the matrix of second derivatives is positive definite at the solution. Now it is required that $t \leq y \leq 1$ and $r \geq 1$, so that (10) and (11) apply on $S_1 \cap T_1$. By continuity, on $S_1 \cap \bar{T}_1$ take $y_o = \frac{\varphi}{c_s}$ but $r_o = -1$. On S_2, take $y_o = t$, and substitute in (9) to obtain

$$\frac{t}{2 + r_o t} = \sqrt{\frac{c_s}{2}}$$

and thus

$$r_o = \sqrt{\frac{2}{c_s}} - \frac{2}{t} ,$$

which applies on T_2. Take $r_o = -1$ on $S_2 \cap \bar{T}_2$. Finally, on S_3 take $y_o = 1$, which gives

$$r_o = \sqrt{\frac{2}{c_s}} - 2 ,$$

and which applies on T_3. Take $r_o = -1$ on $S_3 \cap \bar{T}_3$. The minimum values of $L^*(y,r)$ are obtained by substitution.